Chapter 2

ISSUES OF GENERALIZATION, DISAGGREGATION, AND COMPARISON

In this chapter we report the results of additional analyses of the NWBS data. Three issues are addressed. The first is the estimation of a valuation function which can be used to estimate the benefits of specific changes in water quality. The second is the use of this valuation function for the purpose of providing estimates of the benefits of regional changes in water quality. The third is an examination of the average per recreation day values for boating, fishing, and swimming implied by our NWBS data.

ESTIMATING MARGINAL FRESHWATER QUALITY BENEFITS USING A VALUATION FUNCTION

Valuation Function

A total value/bid curve for a single economic agent (such as a household) can be specified which depicts the Hicksian compensated variation, $TOTWTP_i$. as a function of the level of water quality, q_i, household income, Y_o, water based recreational use, W_r, and environmental attitudes, A_e, for a given base water quality level, q_o.

$$TOTWTP_i = f(q_i, Y_o, W_r, A_e \mid q_o) \qquad (1)$$

Differentiating this bid curve with respect to q_i, $\partial TOTWTP_i/\partial q_i$, yields the inverse Hicksian compensated demand curve.

To empirically estimate equation (1) we stacked the observations from each of the water quality levels, which the respondents were asked to value, and used the common log-log function form, which appears to fit better than a linear form in our regression. The results can be summarized by,

$$\log(TOTWTP_i) = \underset{(-7.79)}{-1.703} + \underset{(13.28)}{1.258}*\log(q_i) + \underset{(19.35)}{1.107}*\log(Y_o) + \underset{(3.70)}{.180}*W_r + \underset{(6.52)}{.362}*A_e, \quad (2)$$

where q_i is the rung (e.g. boatable) on the water quality ladder (figure 1.2) being valued, Y_o is the current annual household income in thousands of dollars, W_r is operationalized as a dummy variable for whether or not any member of the household engaged in freshwater boating, fishing, or swimming activity during the previous year, and A_e is operationalized as a dummy variable for whether or not the respondent regarded a national goal "of protecting nature and controlling pollution" as "very important". The t-statistics given in the parentheses are based on the heteroskedasticity consistent covariance matrix proposed by White (1980). The coefficients in this equation are all reasonable in terms of sign and magnitude and are all quite significant. The adjusted R^2 for this equation, .31, is substantial considering the small number of variables in the equation and the use of very heterogeneous cross-sectional survey data.

The particular specification of the recreation variable, W_r, was chosen because it can be obtained at a state, and sometimes at a sub-state level, from recreation surveys. The environmental attitude variable, A_e, differs from the one used in our original report (Mitchell and Carson, 1984) which measured a belief that the United States is spending too little on water pollution control. That variable posed the possibility that its predictive power was an artifact of its apparent similarity to the WTP questions. The current variable comes from a question about how important the respondent believes the much more general goal of protecting nature and controlling pollution is. It's presence. at a statistically significant level, confirms our belief that environmental attitudes are a significant predictor of WTP for water quality.

For the sample on which this regression was based, described in the preceeding chapter, $Y_o = 24.22$, $W_r = .59$, and $A_e = .65$. While there is a fair amount of variation in A_e across individuals, we find no statistically significant variation in A_e across the four census geographic regions for which we have representative subsamples.[1] $\log(TOTWTP_i)$ was set equal to zero if $TOTWTP_i$ equaled zero.[2]

Note that taking the exponents of the predicted values from this equation results in estimates of the conditional median WTP rather than mean WTP (see Goldberger, 1968; Stynes, Peterson, and Rosenthal, 1986).[3] The presence of heteroscedasticity causes poor prediction of mean WTP for larger values of q_i and Y_o. This is because the variance estimate is important in predicting conditional means from the log-log equation which implicitly assumes that they are distributed i.i.d. normal after taking the log of the error terms in order to perform the retransformation.

For predicting conditional mean WTP, we find the same basic relationship implied in equation (2) -- but with an additive rather than multiplicative error term (Goldfeld and Quant, 1972) -- works well and can be estimated using nonlinear regression techniques. This yields

$$TOTWTP_i = EXP[.413 + .819*\log(q_i) + .959*\log(Y_o) + .207*W_r + .460*A_e], \quad (3)$$

$$(1.66) \quad (9.20) \quad (22.30) \quad (3.23) \quad (6.47)$$

---

1. The Census Bureau defines two sets of geographic regions, one consisting of nine and the other a more aggregated set of four regions. Our 1980 study used a larger sample and a stratified sampling plan which allowed us to generalize to the nine region set.

2. This is equivalent to setting $TOTWTP_i$ equal to one if it was zero, and has little effect on predicted values since both the mean and median WTP amounts are not close to zero.

3. The conditional median from equation (2) is given by $EXP(X\beta)$ while the conditional mean is given by $EXP(X\beta + \sigma^2/2)$ if $\log(TOTWTP_i)$ is assumed to be normally distributed.

where the asymptotic t-statistics are in parentheses, n=1599, and $\bar{R}^2=.27$.

The value of an incremental water quality change from q_r to q_s, $INCVAL_{(r,s)}$, using equation (3) can be estimated by evaluating,

$$\int_{q_r}^{q_s} EXP[-.413 + -.819 + \log(q_i) + .959*\log(Y_o) + .207*W_r + .460*A_e]\, dq \qquad (4)$$

This can be done by substituting appropriate values for Y_o, W_r and A_e into equation (3) and solving it twice, once for q_r and once for q_s. Equation (4) can now be restated as,

$$\hat{INCVAL}_{(r,s)} = \hat{TOTWTP}_s - \hat{TOTWTP}_r. \qquad (5)$$

Using the mean values of Y_o, W_r, and A_e from our survey to solve equation for a change from boatable (q_r=2.5) to fishable (q_s=5) yields an estimate of $INCVAL_{(r,s)}$ of $79. This estimate is almost identical to our actual observed in-sample value of $80. Use of this valuation function to estimate the marginal benefits of other water quality changes is straightforward.[4]

Valuing Subnational Water Quality Changes

A more ambitious use of this valuation function is to value the marginal benefits of water quality changes in a local area where less than all of the water bodies in an area are valued. To do this three unknown multiplier functions are needed. The population multiplier (POPM) is simply the size of the population whose values for freshwater quality will be aggregated in order to obtain the local estimate. The second multiplier, percent local water benefits, PLWBM, describes what percent of the WTP amount described by the

4. The water quality change of interest must be translated into initial and subsequent levels on the RFF water quality ladder. See appendix B for details of how to perform this mapping.

national valuation function is intended to apply to the entire subnational geographic area under consideration. The third multiplier is area water change, LWACHM, which describes how WTP changes as a function of the quantity of water in the local area being changed. Using these multipliers allows household aggregate WTP for a change from q_r to q_s in a local area, $AGINCWTP_{(r,s)}$, to be defined as,

$$AGINCWTP_{(r,s)} = POPM * PLWBM * LWACHM * \int_{q_r}^{q_s} f(q_i, Y_o, W_r, A_e \mid q_o)\, dq, \qquad (6)$$

where the expression behind the integral in equation (6) is taken from equation (1). Operationalizing equation (6) will require values for POPM, PLWBM, AND LUACHM. We now turn to a discussion of the problems involved in arriving at defensible decisions about the magnitude of these multipliers.

POPM. There are two basic approaches to defining the relevant population, POPM where the water improvement to be valued involves water bodies which constitute a portion of a state's total freshwater. The first is to have everyone in the state pay for the improvement. This approach would be most appropriate when both people and water bodies of given quality levels are uniformly distributed across the state, and is formally consistent with the data gathered in the NWBS where we used "state" to refer to their local area when we asked people to apportion their national WTP amounts between their local area and the rest of the nation. However, to the extent that there is a disproportionately large percent of the state's population near the relevant water bodies and the water bodies themselves constitute a disproportionately small proportion of the state's water, this approach will tend to underestimate the true benefits. The converse will be true if there is a low percentage of the state's population and a high percentage of the state's water in the affected "local" area.

The second approach is to restrict POPM to those who reside within a specified sub-state area near the relevant water bodies. This approach assumes that people are more concerned about the water quality in their area than water quality in other parts of the state and is perhaps more intuitively plausible than using the state as the population. It is also consistent with Sutherland and Walsh’s (1985) finding for a Montana water-based recreation site that the WTP amounts for the site declined with their respondents’ distance from the site and with Greenley, Walsh, and Young’s (1982) finding, in their study of water quality benefits for the South Platte River Basin in Colorado, that almost all of their local respondents WTP for water benefits for state improvements were given to the South Platte.

PLWBM. The second multiplier is the percent of the national amount to be applied to the regional situation. Our survey provides some useful information about this multiplier. After obtaining their WTP amounts for the several water quality changes, as noted we asked the respondents to apportion their WTP amounts between state water quality improvements and improvements in the rest of the country. Respondents, on average, wanted 2/3 of their WTP amounts to go to improve local water quality. The median proportion was almost the same, We used “state” in this question because our pretests revealed that local area meant different things to different people whereas the term state was universally understood. We also observed during the pretesting, however, that the “local” and “state” percentages did not vary too much from each other, particularly if the respondent tended to define the local area as being regional in scope. Gramlich’s study of the Charles River in the Boston area is also consistent with our estimate of the appropriate magnitude of the PLWBM. Gramlich asked people living in the Charles River Basin what they were willing to pay for the Charles River and for all waterbodies in the United States. The ratio of the WTP amounts for swimmable quality water in the Charles River to the WTP amounts for all water bodies was .55 and there are several other water bodies in the Boston area.

This information and the POPM decision offers some guidance as to the appropriate PLWBM coefficient as it suggests that this multiplier should be no higher than .7 for a statewide water improvement. When a smaller area is involved it may be reasonable to consider smaller values. If the area is not too small and is self-contained, such as a river basin, it could be argued that .5 is a reasonable lower bound for the multiplier. However, if the area is smaller than the area where people perceived most of their freshwater substitutes to lie, a much smaller value of PLWBM may be possible. This suggests that use of the valuation function is best restricted to river basins or portions of river basins which are relatively self-contained.

<u>LWACHM</u>. This multiplier describes how $AGINLWTP_{(r,s)}$ changes as a function of the change in water quality being contemplated relative to the change represented by the NWBS for the local area. LWACHM can be defined as

$$LWACHM = h \left[\frac{\text{Perceived quantity of area water affected by proposed (BOAT,FISH) policy}}{\text{Perceived quantity of area water affected by (BOAT,FISH) change in NWBS.}} \right]$$

There are three important things to note about this function. First, area is defined by the choice of POPM. If POPM is the state, for example, then the local area water is the freshwater in the entire state. If POPM is that of several counties, then the freshwater in those counties will be denominator in the LWACM expression.

Second, it is based on the respondents' perceptions about the changes rather than the actual water conditions. Beyond the standard implication that it is perceived changes rather than actual changes that make agents better off, there is the further implication that we need to know how respondents conceive of local water. Do people perceive the relative amount of water being changed in terms of surface acres, bank feet, volume, or in some other fashion? In the absence of information about what people have in their minds when they think of the freshwater in a reasonably sizable local area, we will assume that people's conception of freshwater in their area is dominanted by the larger, well known, water bodies in the area. If this is indeed true, then a potentially useful objective measure of the amount of water in an area is the definition used by the Census Bureau which is known as "census water." This is defined as the surface acres of water in all lakes over 40 acres in size and in all streams over 1/8 of a mile wide.

Third, this function is potentially nonlinear. Only in the special case where average valuation is equal to marginal valuation of water quality improvements over the change considered will LWACHM be a linear function of the relative percent of water involved in the change. As it turns out, the data we presented in chapter 1 for one of the two partial improvements we asked our respondents to value is relevant to marginal changes in the percent of improved water. This partial improvement was valued by asking the respondents how much improving 95 percent of the water from boatable to fishable was worth to them compared to their values for improving 99 percent of the water. Our findings indicate slightly increasing marginal valuation for the last remaining percentage improvements. The effect is not large, however, and is confounded to some degree by our statement that the water not improved would be in populated urban areas. And our pretests which inquired about the difference between 99 percent and 100 percent improvements, suggest that respondents do not regard moving the last small portion of freshwater to a higher level as a highly symbolic and therefore highly valued act. This would rule out the extreme increasing marginal valuation of the type found by Brookshire, Schulze, and Thayer (1985) for air visibility in the Grand Canyon. We conclude on the basis of this admittedly scanty evidence that marginal and average may be approximately equal over a fairly large (and the most relevant for policy purposes) range.

CALIBRATION WITH THE MONONGAHELA DATA

We have emphasized the problems involved in applying our valuation function to local area water quality improvements. The study design for the NWBS was oriented towards a national water quality policy whose goals are to raise all of the nation’s waterbodies to the boatable-fishable-swimmable level. The relationship between intermediate improvements in a particular river and what the respondents in our sample were valuing is necessarily uncertain. In order to assess the validity of our approach, we need to compare estimates for a local area improvement based on our valuation function with independently measured benefit estimates for the same improvement. The Desvousges, Smith, and McGivney (1983) (DSM) CV study of water quality in the Monongahela River Basin offers such an opportunity. DSM's study was quite similar in design to NWBS and thus offers the opportunity to calibrate our valuation function for local use and to make some firmer recommendations on appropriate multiplier definitions and values.

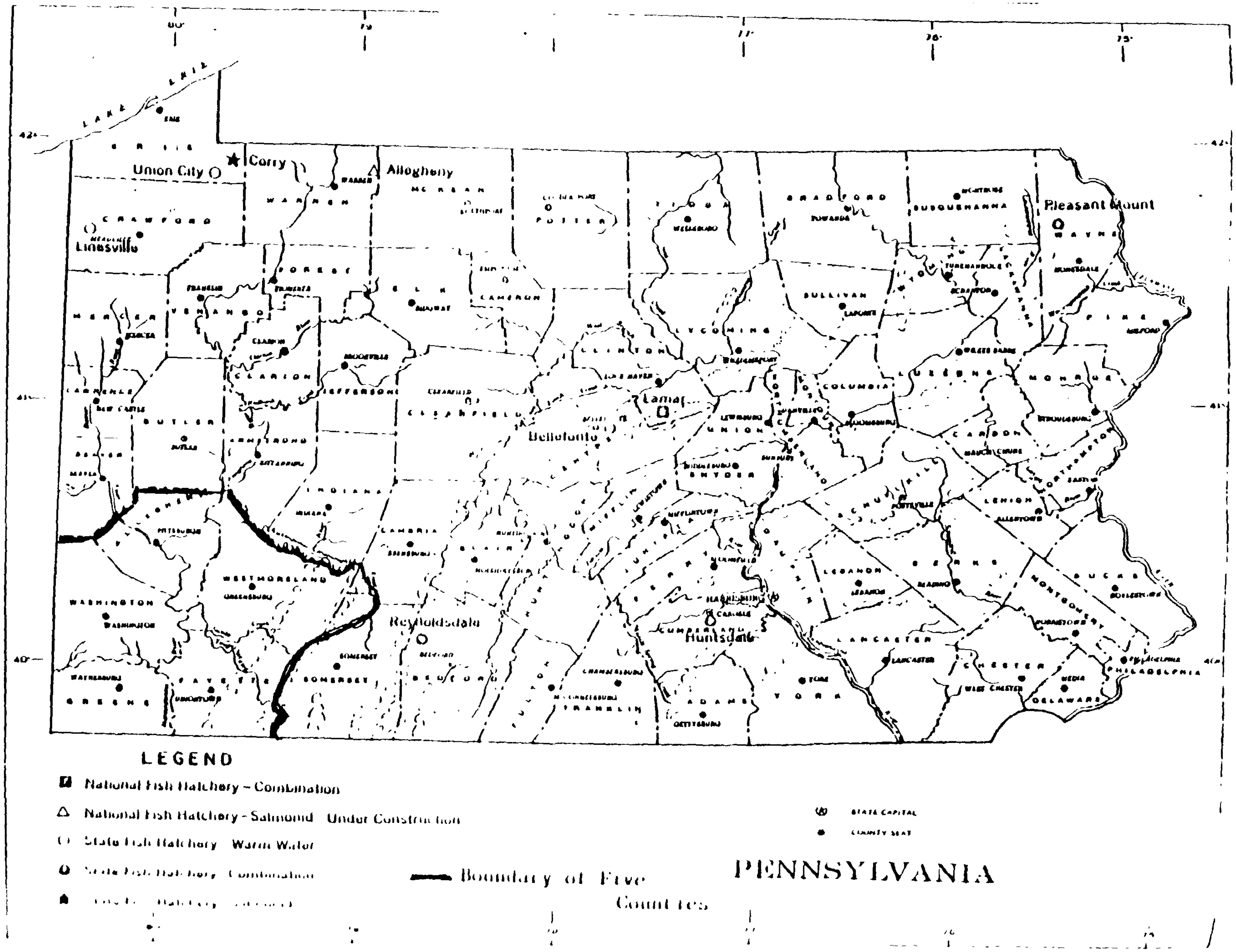
DSM conducted a CV study in 1982 to value water quality changes in the Pennsylvania portion of the Monongahela River and its tributaries. The Monongahela river runs from the West Virginia border north to Pittsburgh. At

Pittsburgh the Monongahela is joined by the Allegheny River, whose watershed is northeast of Pittsburgh, to form the Ohio River which runs west. DSM used personal interviews of a stratified random sample of residents in the five Pennsylvania counties through which the Monogahela flows. Their study was designed to test the effect of different elicitation formats, so the sample was divided into equivalent subsamples, each of which received a different elicitation method.

Of particular importance to the comparison we report here is the fact that DSM adopted our water quality ladder; the use of the boatable, fishable, swimmable quality levels; and our payment vehicle of higher taxes and prices in their scenario. This greatly assists making the comparisons between our values for a change from boatable to fishable for national water quality and the ones they obtained for the same change in the Monongahela.

DSM defined the water bodies they wished their respondents to value as follows. First they asked their respondents to look at a map of the area and a list of 29 recreational sites, fifteen of which were on the Monongahela River. The remainder were on the Allegheny River. A number of these sites involve portions of the rivers which are lakes created by Corps of Engineers dams. After obtaining information about the respondents' use of these sites for recreation, the respondents were told to restricted their values only to the Monongahela River and its major tributaries in Pennsylvania and to assume that the water quality at the other sites on the list (the 14 Allegheny sites) "stays the same as it is now." They specifically invoked both use and existence values for the river. Then they specified their interest in the quality of the river as a whole -- in other words, the minimum quality level. Finally they asked their respondents to value the same series of improvements as we did in the NWBS, (1) to maintain boatable, (2) to go from boatable to fishable, and (3) to go from from fishable to swimmable.

Figure 2-1 MAP OF PENNSYLVANIA SHOWING FIVE COUNTY MONONGAHELA RIVER AREA



As shown in table 2-1, DSM found a $INCVAL_{(BOAT,FISH)}$ for the Monongahela which ranged from $15.9 to $36.9 per year. The low value was obtained in a bidding game using a $25 starting point and the high value was obtained in a bidding game with a $125 starting point. Their payment card treatment, which most closely parallels the elicitation format used in NWBS, found a mean WTP of $29.30. Using $\bar{Y}_o$=$20.88 and $\bar{W}_r$=.56 from DSM's study and the national value of .65 for $\bar{A}_e$, and evaluating equation (4) for the boatable to fishable change gives us an estimated $INCVAL_{(BOAT,FISH)}$ of $68.29 for a household's (with these characteristics), willingness-to-pay for an incremental national change of this magnitude.

Table 2-2 presents basic information about the five county area included in DSM's sampling plan and figure 2-1 shows its location in Pennsylvania. According to the 1980 census, the population in the five counties consists of 827,536 households while the state population consists of 4,219,606 households. It would appear, from reading DSM's scenario (including the visual aids), that their respondents may have believed that approximately 50% of the freshwater in the five county area would be affected by the changes they were asked to value in Monongahela basin water quality.[5]

---

5. Respondents in their survey were handed a card which divided the water in the area approximately equally into the Allegheny and Monongahela river systems.

Table 2-1 DSM’s Contingent Valuation Estimates of the Value of a Change from Boatable to Fishable Water Quality in the Monongahela River

Method	N	Mean Annual WTP Per household (Standard Deviation)	Mean Annual Aggregate WTP (Range)
Iterative bidding, $25 starting point	58	$15.90 ($15.50)	$13,158,000 (9,856,000 to 16,460,000)
Direct question	51	17.60 (32.10)	14,565,000 (7,282,000 to 21,847,000)
Payment card	54	29.30 (49.30)	24,247,000 (13,628,000 to 35,114,000)
Iterative bidding, $125 starting point	48	36.90 (49.50)	$30,536,000 (18,951,000 to 42,122,000)

Desvousges, Smith, and McGivney (1983. p. 4-32). Protest bids and outliers are excluded from these data.

If an objective measure is desired, we can use the previously mentioned Census Bureau definition of water area which includes ponds and lakes over 40 acres and streams and canals at least 1/8 of a mile wide (U.S. Bureau of the Census, 1967).[6] Based on conservative estimates made from very large county water maps, a low estimate of the Monongahela River system’s percent of the census water in DSM’s 5 county area is [illegible]. This estimate excludes the Ohio

6. For a useful compilation of definitions of state recreational water availability see Dyson (1984).

River in Allegheny County. The per county estimates are given in table 2-3. If the Ohio River in Allegheny County is included, the Monongahela River system would constitute roughly 50 percent of the census water in the five county area. On a state basis, as shown in table 2-3, the Monongahela constitutes 1.5 percent if the 24 percent five county figure is used and 3.3 percent, if 50 percent of the census water in the five counties is considered to be affected.

Table 2-2 Census Water and Population Data for Five Counties and the State of Pennsylvania

	Five Counties	Pennsylvania	Five Counties as % of Pa.
Total census water (thousand surface acres)	4.02	269.5	1.5
Household Population (thousands)	828	4219	20.0

Table 2-3 Census Water Estimates for Monogahela River and for Five Pennsylvania Counties

County	Monogahela River[a]	Total for County[b]	Monogahela Percent
	(1,000 acres)		
Allegheny	1.20	8.6	14%
Fayette	.99	3.0	33
Greene	.54	.6	90
Washington	1.05	2.1	50
Westmoreland	.24	2.4	10
Total	4.02	16.7	24

SOURCES

a. Monongahela area estimate based on hand measurement made on maps for each county contained in appendix to Green International, Inc. (1979). In making these estimates it was assumed that the Monogahela is 1/8 of a mile wide throughout its length and that the Ohio and Allegheny Rivers, are 1/4 mile wide. Where rivers run along a county line, half of the water surface in the river is accounted to each county.

b. Census water surface area measurement from 1982 National River Inventory table for Pennsylvania (July 1984).

We are now able to use equation (6) to estimate water benefit values for the Monongahela under a variety of assumptions as shown in Table 2-4. The improvement we will examine is from boatable to fishable quality. This is the most comparable improvement to use because the present level of the Allegheny River, according to DSM, is fishable. In the case of each set of assumptions, the INCVAL is constant at $68.29. The first six estimates are all based on defining POPM as the number of 1980 census households in the five county area in Pennsylvania through which the Monongahela River runs and which constituted the population for DSM's CV survey. Two levels of PLWBM, the proportion of the national WTP amount to be applied to the local area, are used. The first, 67 percent, seems to us to be the upper bound for this value and the other, 50 percent, the lower bound.

Three levels of the multiplier for how the value changes as a function of the change in water quality being contemplated relative to the change represented by the NWBS for the local area, LWACHM, are provided as this is the most problematic multiplier to specify. A value of 1 is the maximum plausible value. Recall that people in DSM's study were told to assume that the Allegheny River would remain at its present level of quality irrespective of the changes the people were asked to value in the Monongahela. The maximum value would be indicated if people saw the Allegheny River system as being of much higher quality than the (BOAT,FISH) change being valued for the Monongahela and if they regarded the Monongahela as constituting the bulk of the water in their local area. Perhaps more likely, based on objective criteria and on the impression conveyed in the DSM survey, people viewed the two river systems as being approximately equal in size and in water quality. This suggests a LWACHM value of less than one and one determined to a large degree by the ratio of the perceived quantity of the area's water affected by the DSM (BOAT,FISH) policy divided by the perceived quantity of water in the area. A value of .5 would be indicated of people perceived the Monongahela as constituting around half the local quantity of water affected by the (BOAT,FISH) change proposed in the NWBS. This is what the DSM scenario appears to imply, but does not directly state. A more conservative value of .24 would be consistent with some of the physical measurements of water quantity depicted in table 2-3.

The state based POPM cases are presented for comparison only, as the even distribution assumption necessary to use the entire state in this manner does not characterize this case. The five county area has 20 percent of the state population and at most only 3.3 percent of its census water.

The annual aggregate WTP amount which we estimate for the five county area on the basis of our national valuation function and the combination of various multipliers ranges between $7 and $37 million dollars. Table 2-1, which summarized the findings obtained by DSM for the boatable to fishable change, gives a lower bound of around $9 million and an upper bound of $42 million. The range for their payment card version is $14 to $35 million dollars. Clearly, our estimates fall well within the range obtained by DSM. More specifically, if we consider the DSM best estimate (based on the payment card mean) to be around $24 million. We can use it to assess, in a rough sense, the validity of our multipliers. On *a priori* grounds, our most plausible set of multipliers are POPM = 827,536, PLWBM = 0.67, and LWACHM = 0.5. This yields an estimate for the boatable to fishable change in the Monongahela River system of approximately $19,000,000. DSM's data would suggest that the 0.5 LWACHM multiplier is to be preferred to the .24 multiplier, and that setting LWACHM equal to approximately .65 would result in a close to perfect fit between the two numbers.[7] This implies that LWACHM may be a slightly increasing function of the affected water ratio or that the public places a higher premium on brining the lowest quality water up to standard. This is an empiricial question and could profitably be the focus of future research.

7. DSM's data also provide further support for our earlier contention that the state-based POPM approach was likely to be inappropriate.

Table 2-4 Estimation of the Value of a Change from Boatable to Fishable Water Quality in the Monongahela Derived by the Valuation Function Under Various Assumptions

POPM	PLWBM	LWACHM	MARVAL	Annual Aggregate WTP (Per Household)
	0.67	1.0	$68.29	$37,863,000 ($45.75)
FIVE COUNTY	0.67	0.5	"	18,932,000 (22.88)
(827,536 households)	0.67	0.24	"	9,087,000 (10.98)
	0.5	1.0	"	28,256,000 (34.15)
	0.5	0.5	"	14,128,000 (17.07)
	0.5	0.24	"	6,782,000 (8.20)
	0.67	0.068*	"	13,214,000 (3.13)
PENNSYLVANIA 4,219,606 households	0.67	0.033+	"	6,371,000 (1.51)
	0.67	0.0156@	"	3,012,000 (.71)

* Based on $\frac{\text{Total Census Water in 5 Counties}}{\text{State Census Water}}$

\+ Based on $\frac{\text{Total Census Water in 5 Counties}}{\text{State Hatchery Fish Need Estimate}}$

@ Based on $\frac{\text{One Half of Census Water in 5 Counties}}{\text{State Hatchery Fish Need Estimate}}$

The State Hatchery Fish Need Estimate is an estimate, made in 1965, of the acreage of water, excluding the segment of Lake Erie bordered by Pennsylvania, which can be utilized for recreational fishing. This is a lower estimate of the freshwater available for recreational purposes.

Conclusion

The valuation function which we have obtained from our national water quality survey allows the estimation of marginal changes in water quality using the mapping provided by the RFF water quality ladder. It is based on a fairly simple, theoretically plausible, model which provides a good fit to the cross-sectional NWBS data.

The outcome of our attempt to use the valuation function from our national study to estimate the benefits of the boatable to fishable water quality improvement in the Monongahela River that DSM valued is quite positive. our aggregate estimates, based on reasonably plausible multipliers resulted in to aggregate estimates that coincided quite closely with those estimated on the basis of DSM's WTP amounts. This is important evidence for the convergent validity[8] of our NWBS and adds to the credibility of both our study and theirs.

Our success in predicting DSM's results also lends credibility to the use of our valuation function to value local improvements. We wish to reemphasize, however, the problematic nature of this enterprise and to urge appropriate caution in its use for this type of purpose. The valuation function requires the use of the several multipliers and the grounds for choosing the particular multiplier values are still largely arbitrary.

---

8. Convergent is one of several types of validity (Mitchell and Carson, forthcoming, chapter 8). It concerns the correspondence between a measure and other measures of the same theoretical construct. To the extent that a correlation exists -- the measures converge -- the validity of <u>each</u> measure is confirmed.

A NOTE ON RECREATION DAY VALUES

We conclude this chapter by examining the WTP amounts given by respondent households who use freshwater for recreation in additional detail. The NWBS obtained considerable information on the water-based recreational activities of every member in the respondents' households. Although the NWBS was not designed to obtain activity day estimates, we can calculate what are probably best described as "pseudo" unit day recreation values. Our purpose in doing so is to compare them with the available values in the recreation literature. A rough correspondence will be taken as further evidence for our study's convergent validity.

Our procedure is to subset our data set so we can examine the following pairs of respondents: (1) those whose households contain one or more members who have boated on a freshwater body at least once during the previous year vs. those whose households do not contain a boater, (2) those with one or more fishermen vs. those without a fisherman, and (3) those with one or more swimmers vs. those without any swimmers.

One possible approach would be to regress WTP for the boatable quality level on the number of household recreation days plus other household characteristics such as income and to repeat the procedure for fishable and swimmable WTP amounts.[9] This approach does not work because it often results in negative marginal values for days similar to those found by Vaughan et al. (1985). There appear to be two reasons any this happens in our data. First, there is a fair amount of multicollinearity between low income and the number

9. There are two possible measures of WTP for fishermen. One would be the amount to go from below boatable to fishable, the other would just be the boatable to fishable increment. The former would be based on the assumption that those who wish to fish might be willing to pay for the lower quality level because it raises the water closer to the fishable level.

of days of water based recreation. A number of poor folks do a lot of fishing. Second, and related to the first factor, we had a small number of households who engaged in a very large number of recreation days but had a $0 or small WTP.

We abandoned the marginal approach as unworkable in favor of what might be called the average approach. Our procedure in making the following estimates was simply to divide the WTP amount for the relevant incremental improvement by the number of recreational days at that level. The means, medians, and ranges as well as information on the number of recreation days of each type is given in table 2-5. Our procedure is not only simplistic, it also suffers from many problems such as the fact that our respondents undoubtedly engaged in multiple activities on the same day. Nevertheless, our values are clearly within the range of those reported by Loomis and Sorg in their critical summary of empirical estimates of the values of recreation. We take this correspondence as an additional, albeit very rough, indication of the convergent validity of our national benefit estimates for freshwater quality improvements.

Table 2-5 Average WTP Amount Per Waterbased Recreation Day* for NWBS and Loomis and Sorg for Different Types of Water Based Recreation

	Boating	Fishing	Swimming
NWBS			
No. of trips			
Mean	24	24	48
Median	10	12	20
Range	1-202	1-354	1-710
Increment WTP			
Mean	$16.30	$11.97	$13.11
(SEM)	(2.43)	(1.32)	(2.87)
Median	$4.17	$2.50	$1.22
Range	$0-313	0-250	0-500
N	195	225	228

	Boating and Mixed	Cold and Warm Water Fishing	Watersports
Loomis and Sorg**	$6-34	$8-37	$10-27

* Any day where respondent or member in respondent's household engaged in boating, fishing or swimming in any freshwater body. No effort was made to determine the number of hours in which the person engaged in this activity.

** (1982) Activity day values based on review of numerous studies. Values are updated to 1982 dollars.

Chapter 3

THE USE OF SIMULATED POLITICAL MARKETS TO VALUE PUBLIC GOODS

An important question for CV studies of pure public goods, such as national water quality is the validity of the estimates. In the previous chapter we examined some evidence for the convergent validity of the NWBS estimates. In this chapter we report the findings of a study we conducted to explore this issue in a somewhat more indirect fashion. The study was a telephone CV survey conducted in California to value a program to improve the quality of the state's freshwater bodies. This program involved a bond issue which was presented to the state electorate as a referendum in the November 1985 election. Our survey was conducted six. weeks prior to the election. In it we: (1) simulated the referendum in an attempt to predict the vote and (2) used a new method to obtain WTP amounts for different tax prices. An important premise on which the study was based is that the appropriate model for CV studies is political markets rather than consumer goods markets.

In what follows, we first discuss the rationale for a political market model. We then propose a new approach to CV studies in which a referendum is simulated at different tax prices. This method uses a take-it-or-leave-it elicitation method with a single iteration, a procedure which is particular!? appropriate for telephone surveys. We next present the findings from our application of this method in a survey designed to predict the outcome of the California vote on Proposition 25 in the November 1985 election. We conclude with a discussion of the implications of this study for the contingent valuation method in general and our NWBS in particular.

THE REFERENDUM MARKET MODEL FOR CV STUDIES

Until recently, it was assumed that the market which CV surveys should emulate is the private goods consumer market. In its pure form,[1] such a model embodies the notion of a consumer with realized tastes whose purchase decisions are based on a full understanding of the available alternatives based on prolonged experience in the market (e.g., Bishop and Heberlein, 1979). An alternative framework which has received support from some CV researchers,[2] is provided by the political market model. There is a considerable body of theory developed by economists and political scientists (e.g., Deacon and Shapiro, 1975; Enelow and Hinich, 1984; Bergstrom, Rubenfeld, and Shapiro, 1982; Langkford, 1985) on these markets. The form most relevant to the CV method is the referendum, where the voter is faced with a one-time (or at best with a very infrequent) choice on a predetermined policy package to which she must vote yes or no.[3] Here the behavior to be predicted by a CV study would be how informed voters would actually vote if the proposition to provide the amenity was actually on the ballot. The voting decision suggests a more complex, and some (e.g.. Morgan, 1978) would say a more realistic, model of decision making than the one implied by the idealized private goods market model. Instead of assuming that people express preexisting well-realized preferences, this model assumes that people make

---

1. Market researchers (Bettman, 1979) have long recognized that many purchases are infequently made and that the information people gather prior to purchase decisions differs greatly depending on the purchase situation, the type of good, and their past experience.

2. See Ridker (1967); Cummings, Cox. and Freeman (1984); Lareau and Ray (1985); Randall (1986b); and Cummings et al. (1986).

3. In virtually all states voters may vote on binding propositions placed on the ballot by state legislatures, and come states provide for advisory referendums as well (Magleby, 1984: 1).

choices which are influenced by multiple motives, by contextual factors, and by less than perfect information.

Several aspects of referenda recommend them as an appropriate model for CV surveys. First, referenda are actually used as a mechanism to enable citizens to make binding decisions about the provision of public goods such as a new school building or a water pollution control program to be financed by a bond issue. Second, the voter's decision in a referendum has clear economic implications for his household which will have to bear its share of any cost implied by the proposal if it passes. Third, it is a plausible choice framework for respondents. They are likely to be familiar with its method of operation and its use in the political system. Finally, a referenda model lends itself well to the survey setting. A ballot, after all, is similar to a multiple choice questionnaire, and political polling to predict election outcomes is a well recognized feature of public life.

SIMULATING REFERENDA[4]

Economists are interested in the demand curve/function for a public good as a function of price, a relationship which is often summarized in terms of an elasticity estimate. The appropriate elasticity measure in the case of a

4. We note here the work which has been the most influential in developing this synthesis. From the public finance literature, Bergstrom, Rubinfeld. and Shapiro (1982) for their use of survey data as a surrogate for observing individual votes in a referendum. for their use of more than one simple yes/no question, and for their ability to estimate fairly precisely a price elasticity. From the contingent valuation literature, the work of Cummings (Cummings, Cox, and Freeman. 1984 Cummings et. al., 1986) which advocated moving toward some type of spring mechanism to make the willingness-to-pay decision more concrete. From the biological statistics literature Cox (1970) and Finney (1978). We also note a particular intellectual debt to Bishop and Heberlein (1979; 1980) who used a discrete choice format to elicit WTP amounts in a different context and for a different purpose.

referendum describes how the percent of voters willing to vote for a given proposition changes with a percentage change in tax price. Letting V stand for the percent of the voters who are willing to vote for the referendum and T for the tax price, we can define the tax price elasticity of voter approval as,

$$\varepsilon_v = \frac{\partial \ln(V_p)}{\partial \ln(T_p)} , \qquad (1)$$

where ln represents the natural logarithm.[5] This definition is somewhat unusual since it is defined over the population of voters rather than an individual or median voter.[6]

In what follows we propose a method of estimating the demand for a public good which is a synthesis of several research approaches: the public finance tradition of examining referendum results, the contingent valuation tradition of asking people directly about their willingness to pay, and, in particular. the biological tradition of estimating dose-response relationships.

The Response to a Stimulus: Referenda Voting and Bioassays

Even though the term elasticity is infrequently used in the bioassay literature, the researcher doing a bioassay is attempting to measure something directly analogous to a demand curve posed in terms of the voting population and a tax price.[7] In a classical bioassay, a large number of test specimens

---

5. T_p can be replaced with a continuous function which describes different tax prices for different members of society rather than a flat tax price which is assumed for convenience here.

6. See Peterson (1975) for an example of the latter approach.

7. Finney (1978 and earlier editions) is a fairly complete source for information on bioassys. The focus of Cox (1970) may be more familiar to economists working in discrete choice situations. See also Finney (1971) and Mead and Curnow (1983).

are randomly assigned to a small number of groups. Each of these groups is then exposed to different levels of the same stimulus (usually a poison or hazardous condition). The number of animals in each group responding to the stimulus (usually by dying) is counted and the percentage responding calculated. These percentages are then plotted against the level of the stimulus and a curve statistically fit to these points. The location estimate of primary interest is usually called the LD50 for the lethal dose where it is estimated exactly 50% of the specimens would die.[8]

To draw out the analogy, tax price is the stimulus and the percent willing to vote in favor is the response. The LD50 point is the minimum response necessary for the referendum to pass which is the highest tax price at which the median voter would vote to pass the referendum. In order to estimate the demand function all we would theoretically have to do is expose some different groups of voters to randomly assigned different tax prices, observe the percent who say they would vote yes on the referendum, and record this percentage and its accompanying tax price.

In actuality, one must go out and survey representative samples of the voters at different tax prices in order to obtain the desired estimate since they are not available in our “lab” for random assignment. The differences between a controlled experiment and random sampling are, however, not as great as one might expect and we will show now it is possible to combine the two.[9] Another difference, which is more of a challenge, is that the random elements in a sample survey are considered to occur through the selection of particular

---

8. The term ED50, for median effective dose, is now frequently used to denote the location where 50% of the specimens would respond since experiments where death is not the expected response are now common.

9. Fienberg and Tanur (1985) develop the parallels between experimental design and sampling techniques.

individuals rather than through random responses to the stimulus by those selected individuals. This difference is important and we will return to it after discussing double sampling.

Sampling: Simple, Stratified, and Double

Because the human population we are interested in is less heterogeneous in its response patterns than are fruit flies and pure strains of white lab rats, large gains in efficiency may be obtained by changing from simple random sampling to stratified random sampling.[10] Consider a variable Y which takes on two values, say the presence (Y=1) or absence (Y=0) of a particular characteristics. We are interested in the estimated percentage, P, of the population with this characteristic.[11]

If we take a simple random sample of size n, $\bar{P}$ is estimated by

$$\bar{p} = 1/n \sum_{i=1}^{n} y_i \qquad (2)$$

If an indicator variable I=1,2 is available and the $VAR(Y|I=1) \neq VAR(Y|I=2)$, it is possible to define two strata based on this indicator variable. Given a maximum available sample size of n and equal costs for sampling units when I=1 and I=2, the optimal (Neyman) allocation of n is proportional to the number of units in each of the i_{th} strata (N_1, N_2) and the variance of each stratum, so that

$$n_i = \frac{N_i\sqrt{p_i q_i}}{N_1\sqrt{p_1 q_1} + N_2\sqrt{p_2 q_2}}, \qquad (3)$$

10. Under simple random sampling each individual in a population has an equal chance of being chosen to be interviewed for the survey. Cochran (1977) is a good source for information on different sampling schemes.

11. We will follow the convention that population values will be denoted by capital letters while sample values and statistics will be denoted by lower case letters.

where $q_i = 1 - p_i$. The stratified sample estimate of P is

$$\bar{p}_{st} = \frac{N_1\bar{p}_1}{N} + \frac{N_2\bar{p}_2}{N} , \tag{4}$$

where $\bar{p}_1$ and $\bar{p}_2$ are calculated within strata in the same manner as the estimator for simple random sampling in EQ. (1). The gain in efficiency of $\bar{1}$st p_{st} over $\bar{p}$ is greatest if all the units for which Y=0 are in one stratum and all the units for which Y=1 are in the other strata. One attempts to choose the variable for stratification that best accomplishes this.[12]

The variable on which the survey stratification is carried out must be visible before the units are chosen. If no suitable variable for stratification is available before undertaking the survey, the double sampling technique proposed by Neyman (1938) offers a useful alternative.[13] Double sampling represents an extremely simple idea. If a good indicator variable is not available, conduct a large survey using a simple random sample; obtain the indicator variable desired; and then reinterview some percentage of the original sample after stratifying on the indicator variable. In many cases, double sampling will be more efficient than a single large simple random sample survey which costs the same amount. Double sampling is likely to be a profitable strategy when the gains from stratification using the indicator variable obtained in the first stage sampling are very large. The percentage of respondents to be reinterviewed in the second phase is a function of the

---

12. We here assume that N, N_1, and N_2 are known.

13. Cochran (1977) provides a discussion of double sampling and relates it to other sampling techniques. Double sampling is used frequently in medical surveys when it is inexpensive to identify a group potentially having desired characteristics but expensive to test for them. See Deming (1977) for an example.

variances in the new strata and the cost of a second stage interview relative to the first stage interviews. Double sampling also allows one to obtain unbiased estimates of population strata sizes N_1 and N_2 when these are unknown, as is often the case.

Efficient Referenda Simulation

In this section, we propose a method of efficiently simulating referenda based on double sampling. First, one draws a simple random sample of voters and asks each of them whether or not they would vote in favor of a specified referendum if its tax price to the voter was x dollars. Having obtained a yes or no response to the x dollar question, this variable is then used as an indicator for stratification. If an individual said “yes” to x dollars then the probability that the same individual in the same interview would say “yes” to any other amount less than x can safely be assumed to be one. Similarly, the probability that an individual who said “no” to x dollars would also say "no" to any amount greater than x dollars is also one. These two statements utilize the weakest variant of revealed preference theory--that if a good is not preferred at one price, it will not be preferred at a higher price and that if it is preferred at one price it will also be preferred at any lover price.[14]

It should be clear that the variance of the responses of those who answered “yes” to the x dollar question will be zero if the subsequent tax

14. We are also using the assumption that if asked the same question twice in the same survey (without any intervening information) a respondent will give the same answer. Biologists, in contrast, assume that lab animals respond to some specified level of a stimulus only in a probabilistic fashion. Note Finney’s (1978) criticism of a method proposed by Dragstedt and Land (1928) which implicitly assumes that if an animal did not respond to a specified dose, it will not respond to any dose lower than that one.

price asked is less than x since these respondents would all answer “yes".[15] From EQ. (3) we can see that if there are a relatively small number of respondents who can be reinterviewed, the optimal allocation of these reinterviews between the two strata (using as the strata indicator variable the respondent’s answer to the x dollar tax price) would be no reinterviews in the stratum in which the responses would have zero variance and all of the available reinterviews in the other stratum. The allocation depends on whether the new tax price was greater or less than x dollars. This scheme can obviously be repeated based on the responses to an initial tax price of x_1 dollars and the next tax price of x_2 dollars.

To estimate a demand curve, we need to know the percent in favor at several different tax prices. Two key questions are: how many different initial tax prices should be used, and how many followup tax prices can be used before the quality of the respondents’ answers deterioriate because of contamination by the starting point and the numbing effect of repetitive questions. There is an obvious tradeoff between, how well we can estimate the percentage acceptance at a particular tax price and the number of different tax prices for which we can obtain estimates given a fixed sample size.[16] The larger the number of tax prices used and the smaller the sample size at each tax price, the less the precision in the estimate of $\bar{p}$ at that point. In most cases, the overall sample size will be limited by the researcher’s budget, and it should be realized there are severe response problems with

---

15. This statement also holds for the variance of the respondents who said “no” to x dollars if the subsequent amount is more than x dollars

16. In cases where the researcher is economizing by adding questions to a multipurpose “omnibus” survey, the survey researcher organization will place constraints on the number of tax prices and the number of “follow up questions/reinterviews”.

asking people a series of very repetitive questions (Mitchell and Carson, forthcoming). Thus we cannot recommend using more than one or, possibly, two followup tax prices.

We can rule out some parts of the demand curve as being essentially uninteresting for our purposes, such as referenda with tax prices so high that they stand no chance of being approved by 50% (or more) of the voters.[17] Often the researcher's interest will lie in the area of a particular proposed tax price for the public good in question.

AN EMPIRICAL TEST

California's frequent statewide referenda offer a useful context for a natural experiment whereby we first use our method of simulating referenda in a survey and then compare our hypothetical results with the actual outcome of a referendum vote (at a single tax price) for the same public good.[18] Proposition 25 ("Clean Water Bond Law of 1984") was selected as our target referendum because of its relevance to the NWBS. This noncontroversial proposition, which was put on the ballot for the November 1984 statewide election by an almost unanimous vote of the state legislature, authorized a 20 year bond issue of 325 million dollars largely for the purpose of constructing sewage treatment plants. Much of the money raised from the sale of the bonds

---

17. This may not always be the case. In particular if the researcher is interested in mean consumer surplus, estimates of the extreme quantiles become very important, and special techniques may be called for. We return to the issue in the concluding section.

18. While California is perhaps unique in the range of topics and frequency of occurrence of referenda, referenda, particularly on local issues, are commonly held throughout the country (Magleby, 1984). The study of California referenda has long been popular with both economists and political scientists. See Wolfinger and Greenstein (1968), Mueller (1969), Deacon and Shapiro (1975). Lutrin and Settle (1975), Lake (1983). Conway and Carson (1984).

would be used as the state's matching contribution for Federal sewage treatment grants.[19] The principal and interest on the bonds would be paid for out of the state's general fund. The cost of Proposition 25 to the average California household is approximately $4 per year for 20 years after interest payments are included. Since the exact distribution of repayment obligation is unspecified from the taxpayers' point of view, we have conveniently assumed a flat per household tax in our simulation. This is approximately correct if sever charges/utility taxes are used to pay off the bonds, somewhat less correct if sales tax revenue is used, and too regressive if the somewhat progressive state income tax is used. A detailed description of Proposition 25 was presented in the California voters pamphlet (Eu and Hamm, 1984) which was mailed to all California voters before the election.[20]

Implementation

We arranged to simulate Proposition 25 at different tax prices on a Field Institute California Poll held in early October 1984. The Field Institute is a not-for-profit organization, and the well-known California Polls are conducted for a group of leading California newspapers and broadcasters as well as a consortium of California Universities. The California Polls are

19. This is the fourth bond issue of this nature put to a referendum vote in California since 1970. The other referenda in 1970, 1974, and 1978 all passed comfortably and authorized a total of 925 million dollars in bonds. Radosevich (1975) provides an analysis of the 1974 referendum from a political science prospective.

20. it was received by the voters after our referendum simulation took place. although our description of the simulated referendum in the survey questions was based on information given to us by the Secretary of State's office which prepared the voter's pamphlet. Sixteen other propositions appeared on the ballot in this election. Proposition 25 received almost no publicity, being overshadowed by the Presidential race and some of the other propositions. It received cursory endorsement from most of the media.

devoted to examining issues of national and state politics. In this particular California Poll, the focus was on the November 1984 general election, providing an excellent and realistic context for our simulation.

The topics we asked about were:[21]

(1) Was the respondent aware of Proposition 25? [AWARE25].

(2) How did the respondent intend to vote on Proposition 25 given the brief description of the referendum which would appear on the ballot in the November election? [VOTEREF]

(3) If the referendum were to cost their household $4 a year for 20 years (the amount implied in the legislative analysis contained in the larger election information pamphlet mailed out by the California Secretary of State's office in mid-October), how would the respondent vote? [VOTE4]

(4) [Depending on the response to the question above] How did the respondent intend to vote on one of three randomly assigned lover amounts (if the vote at $4 was no) or one of six randomly assigned higher amounts (if the vote at $4 was yes)? [VOTE1, VOTE2, VOTE3, VOTE5, VOTE7, VOTE10, VOTE15, VOTE25, VOTE50]. The assignment scheme is discussed in more detail below.

The issues we wish to address here are:

(a) Is the response to VOTEREF independent of VOTE4?

(b) Does providing the respondent with information on the cost of the referendum result in reducing the number of "don't know" responses?

(c) How sensitive is the percent of California citizens who are willing to vote for Proposition 25 to the cost of that referendum (i.e., that is the price elasticity of voter approval at different points)?

(d) What demographic and attitudinal characteristics are associated with the response to VOTE4 (or VOTEREF)?

Our sample consists of one thousand twenty-two respondents were selected by random digit dialing sampling and interviewed by telephone. Assuming that

---

21. We were able to repeat the first two questions one month later on the California Poll taken immediately before the November General Election. There was approximately a 15% increase in awareness of Proposition 25 (from 30%) and almost no change in VOTEREF.

every California household containing one or more registered voters has a telephone (listed or unlisted) and that there was no systematic nonresponse effects, this sample can be treated as if it was a simple random sample.[22]

The VOTEREF and VOTE4 questions are from this sample as are the attitudinal and demographic variables used to answer question (d).

The sampling design for the VOTE1, VOTE2, VOTE3, VOTE5, VOTE7, VOTE10, VOTE15, VOTE25, and VOTE50 questions was based on the six independent replicates making up the Field California Poll and the response to the VOTE4 question. To estimate the sensitivity of the percentage of California voters in favor of Proposition 25 to the tax price, the following experiment within a survey (Fienberg and Tanur, 1985) was conducted. If a respondent answered "no" to VOTE4, then that respondent was asked one[23] of the following: VOTE1 [would you vote for the referendum if the cost were $1 per year]. VOTE2 [cost $2 a year], or VOTE3 [cost $3 a year].[24] From these responses, the percent who would vote in favor of the proposition at each of the dollar amounts can be easily calculated.

---

22. There was a simple stratification between Northern and Southern California and some very minor clustering within telephone area codes. These effects are likely to be small due to the broadness of the strata and low inter-cluster correlations and thus are ignored. A detailed description of the sampling plan for this survey is available from the Field Institute.

23. The assignment depended on which of the six replicants the respondent was in, with two replicants each assigned to VOTE1, VOTE2, and VOTE3.

24. There are really three strata. The third being those who stated don’t know to VOTE4. These respondents are for now considered to be against the referendum at any price. (The original design called for the "don't knows" to be treated as "no," but this was not implemented as intended during the execution of the survey.) Don't know responses to VOTE4 were not asked any additional willingness to vote questions.

Independence of VOTEREF and VOTE4

A political scientist or sociologist would assume that the respondents' answers to VOTEREF and VOTE4 are independent. In an economist's world, though, questions about a person's wants are clearly suspect unless they are tied to the cost of fulfilling those wants. We know from past experience that questions like VOTEREF are good predictors of actual outcomes on referenda [Magleby (1984)].[25]

VOTEREF and VOTE4 each have three possible responses, "yes", "no", and "don't know". A 3 x 3 contingency table can be formed by cross-classifying the responses to these two questions. This table with the actual frequencies. as well as row, column, and cell percentages and marginals is given in table 3-1.

25. Proposition 25 passed in the November election with 73% of the vote. Studies of voting behavior in bond issues indicate that many of our undecided voters are going to vote, and that more than half will vote no (thus preserving the status quo). Splitting the undecideds 50/50 results in 75% in favor, 60/40 against results in 73%, while splitting them 70 - 30 results in 70% in favor. The 95% confidence interval for percentages this region for this size sample is roughly plus or minus 3%.

Table 3-1 California Voters' Intended Vote on Proposition 25 (VOTEREF) and Their Intended Vote Given That The Proposition Would Cost Their Household $4 Per Year (VOTE4)

VOTEREF Frequency Cell Percent Row Percent Column Percent	VOTE4 YES	 NO	 DON'T KNOW	 TOTAL
YES	597 58.4 95.9 71.8	33 3.2 5.1 31.1	13 1.3 2.0 15.3	643 62.9
NO	62 6.1 48.4 7.5	57 5.6 44.5 53.8	9 0.9 7.0 10.6	128 12.5
DONT KNOW	172 16.8 A68.5 20.7	16 1.6 6.4 15.1	63 6.2 25.1 74.1	251 24.6
TOTAL 1 106	83 81.3	 85 10.4	 1022 8.3	 100.0

Denoting the probability that an individual falls into the ith row and the jth column of table I by P_{ij}. we can state our hypothesis formally:

H_1: $P_{ij} = P_{i.}P_{.j}$ for all i and j

K_1: $P_{ij} \neq P_{i.}P_{.j}$ for all i and j.

For large n, a test statistic based on the difference between observed and expected values (under H_1) can be defined (Fienberg, 1977).[26]

26. The X^2 test statistic can be defined as:

$$\sum_{i}\sum_{j} \frac{[\text{observed} - \text{expected}]^2}{\text{expected}}$$

Difference Between Percent of Don’t Know Responses

Providing a respondent with information on how much the referendum will cost him or her if it passes should reduce uncertainty about the referendum, and, as such, should reduce the number of “don’t know” responses.[27] “Don’t knows” on a question like this where the outcome (cleaner water) is perceived as socially desirable are generally of two types: (1) those who are unsure of the cost of passing the referendum, (2) and those who are stating “don’t know instead of stating a “no.” Providing respondents with the cost should remove much although not all) of the uncertainty associated with the referendum’s impact. Most of the remaining responses will be of the “don’t know-means-no”-type. This categorization can also be justified using economic theory in which consumers do not make purchases unless the utility gained from making the purchase is sufficiently above the utility lost at the margin from spending the money. This hypothesis can be stated as:

$$H_2: D_1 = D_2$$

$$K_2: D_1 > D_2,$$

where D_1 is the number of “don’t know” responses to VOTEREF and D_2 is the number of “don’t know” responses to VOTE4. D_1 and D_2 can be thought of as resulting from different binomial distributions and the test of H_2 versus K_2 written in terms of the binomial parameter θ where the estimate of θ is given

27. Don’t knows on a question like this where the outcome (cleaner water) is perceived as socially desirable are generally of two types: (1) those who are unsure of the cost of passing the referendum, and those who are stating “don’t know” instead of stating a “no”. Providing respondents with the cost information should remove much (although not all) of the uncertainty associated with the referendum's impact. Most of the remaining responses will be of the "don't know-means-no" type. This categorization can also be justified using economic theory in which consumers do not make purchases unless the utility gained from making the purchase is sufficiently above the utility lost at the margin from spending the money.

by D/n. For VOTEREF, $\hat{\theta}_1$ = .2456 (standard deviation = .4306), and for VOTE4, $\hat{\theta}_2$ = .0832 (standard deviation = .2762).

Since the binomial distribution has the regenerative property, we can perform an exact test of H_2 versus K_2. The probability of the observed outcome under H_2 is less than .00001, so we reject H_2.[28]

Estimation of the Price Elasticity of Approval

The statistician analyzing a bioassy generally uses some form of quantal response model. Usually a logit (ln[1/(1-p)]) or a probit, the inverse of the standard normal cumulative distribution response function is assumed.[29] In both cases, the maximum likelihood estimates can be obtained using iteratively reweighted least squares and both are members of the family of generalized linear models (McCullagh and Nelder, 1983). We will report only the logit results here. Of more importance is the functional form for tax price. Here both biologists and economists make two typical choices: linear or logarithmic. We will estimate both before assessing the need for other functional forms. At each of the j tax prices (j=1, . . . , 10) there were y_j who indicated a willingness to pay out of a possible n_j individual. (We consider different ways to define the y_j and n_j below.)

The likelihood function for the logistic response model can be written as,

28. As n is large, the normal approximation can also be used. This gives a Z value of 10.15.

29. None of our observations is in the extreme tails, so it will be impossible to distinguish between the two response functions given our sample size. There are generalizations of the logit/probit curves for dose response functions which involve estimating additional parameters if these do not fit the data (Prentice, 1976).

$$l(p,y) = \sum_{j=1}^{10} [y_j \ln(p_j/(1-p_j)) + n_j \ln(1-p_j)] \tag{5}$$

were the y_j are considered to be distributed binomially with index n_j and parameter p_j. We assume that

$$g(p) = \ln(p_j/(1-p_j)) = n_j = \sum_{k=1}^{2} x_{jk}\beta_k, \tag{6}$$

where x_{j1} is a constant term and x_{j2} is either the j_{th} tax price or the natural log of the j_{th} tax price. We can now rewrite the likelihood functions in terms of β.

$$l(\beta,y) = \sum_{j=1}^{10} \sum_{k=1}^{2} y_j x_{jk} \beta_{jk} - \sum_{j=1}^{10} n_j \ln(1 + \exp(\sum_{k=1}^{2} x_{jk}\beta_k)). \tag{7}$$

The logistic response curve may be fit using many standard statistical packages. We used GLIM [Baker and Nelder (1978): McCullagh and Nelder (1983)] because of its convenient facilities for handling the different binomial denominators, n_j. β is estimated by calculating,

$$\hat{\beta} = (X'WX)^{-1}X'Wz, \tag{8}$$

where

$$z_j = \hat{n} + \frac{y - n\hat{p}}{[n_j p_j(1-p_j)]}$$

and W^{-1} = is a diagonal matrix with elements $(\hat{p}_j(1-\hat{p}_j)/n_j)$. The covariance matrix of $\hat{\beta}$ is given by $(X'WX)^{-1}$ since the scale, σ^2, of the binomial distribution is assumed to be one. $\sqrt{\dot{n}}(\hat{\beta}-\beta)$ is distributed $N(0, \dot{n}(X'WX)^{-1})$ plus a bias term of order $O_k(\dot{n}^{-1/2})$, where $\dot{n}$ is the minimum over j of the n_j which goes to zero at the rate of one over the square root of $\dot{n}$.

The number of observations at each price plays a role in the weight matrix, W, as an indicator of the precision with which p_j is estimated; otherwise it only plays a part in estimating the p_j. There are three intuitive definitions we can use for n_j. The first, N*, is the number of respondents who were actually asked the $VOTE_j$ question. The second, N**, is the number of respondents who either explicitly or implicitly answered the $VOTE_j$ question. This is N* plus, for $VOTE_j$ tax prices greater $4, all the respondents not willing to pay $4, and, for $VOTE_j$ tax prices less than S4, all those respondents who were willing to pay S4. Our third definition of n_j, N***, comes from the role n_j, the group sample size under simple random sampling, plays in EQ. (7). The variance of the simple random sample estimator $\bar{p}_j$ is $[\bar{p}_j(1-\bar{p}_j)]/n_j$. Our estimate of p_j is obtained using double sampling. Its variance is a variant of the stratified sampling estimator (Cochran, 1977; 334-335),

$$VAR(\hat{p}_{j(Double)}) = \frac{W_1^2 \bar{p}_{1j}(1-\bar{p}_{1j})}{n_{1j}-1} + \frac{W_2^2 \bar{p}_{2j}(1-\bar{p}_{2j})}{n_{2j}-1}, \tag{9}$$

where p_{1j} and p_{2j} are the simple random sample estimates of the j_{th} p in strata 1 and 2 respectively. W_1 and W_2 are the ratio of the number of observations in strata 1 and 2 to the number of observations in the entire sample (these are the same V j), and n_{1j} and n_{2j} are the number of observations from the two strata sample for the $VOTE_j$ question.[30]

30. A simple numerical example may be helpful here. The total number of interviews was 1022. Of these 831 were willing to pay $4 and 191 were not, so W_1 equals 191/1022 and W_2 equals 831/1022. $\bar{p}_{1j}$ equals zero so the first term of EQ. (9) drops out. We asked 127 $[n_{2(S15)}]$ in strata the VOTE15 question and $\bar{p}_{2(S15)}$ equals .803, so $VAR(p_{S1[Double})$ equals $(831/1022)^2[(.803)(1-.803)/(127-1)]$

Allow us to assume for a moment that n_{1j} and n_{2j} are both greater than 1, in which case the variance contribution from one of the two strata will always be zero since either p_{1j} will be zero or p_{2j} will be one.[31] Ordinarily, one would have to sample from both strata to obtain estimates of p_{1j} and p_{2j}. We have relied on revealed preference theory for the p_j from the strata not sampled thus avoiding the problem of the zero divisor in EQ. (9) for the strata not sampled.[32]

It is possible to set VAR($p_{j\,(Double)}$) equal VAR ($\bar{p}_{j}$ [Simple]) by substituting $p_{j\,(Double)}$ into the formula for the variance of VAR$\bar{p}_{j\,[Simple]}$) so that

$$\mathrm{VAR}(p_{j\,(Double)}) = [\hat{p}_{j\,(Double)}(1 - \hat{p}_{j\,(Double)})]/n_j \qquad (10)$$

and then solve for the n_j associated with the $\bar{p}_{j\,(Simple)}$ with this variance. This is the appropriate n_j to use in the estimation of the demand curve.

Table 3-2 displays p_j , N*, N**, and N***. N*** shows the large increase in efficiency over the number of observations represented by N*. Note how the effect of double sampling is most pronounced for tax prices near $4. For these tax prices, the separation of respondents into two strata of pure types is most successful. It becomes less successful as the strata sampled become more mixed with those who are willing and not willing to pay the specified tax price.

31. If $VOTE_j$ represents a higher tax price than VOTE4, strata 1 will contribute zero to the variance estimate since p_{1j} is always zero, and if $VOTE_j$ represents a lower tax price than VOTE4, strata 2 will contribute zero to the variance estimate since p_{2j} is always one.

32. A Bayesian interpretation can be used if desired.

Table 3-2 Percent in Favor of Proposition 25

Tax Price	Percent in Favor	N*	N**	N***
1	.89	28	926	302
2	.87	38	936	604
3	.85	40	938	1005
4	.81	1022	1022	1022
5	.76	146	337	682
7	.71	144	335	391
10	.72	144	335	421
15	.65	127	318	273
25	.595	127	318	260
50	.48	136	327	210

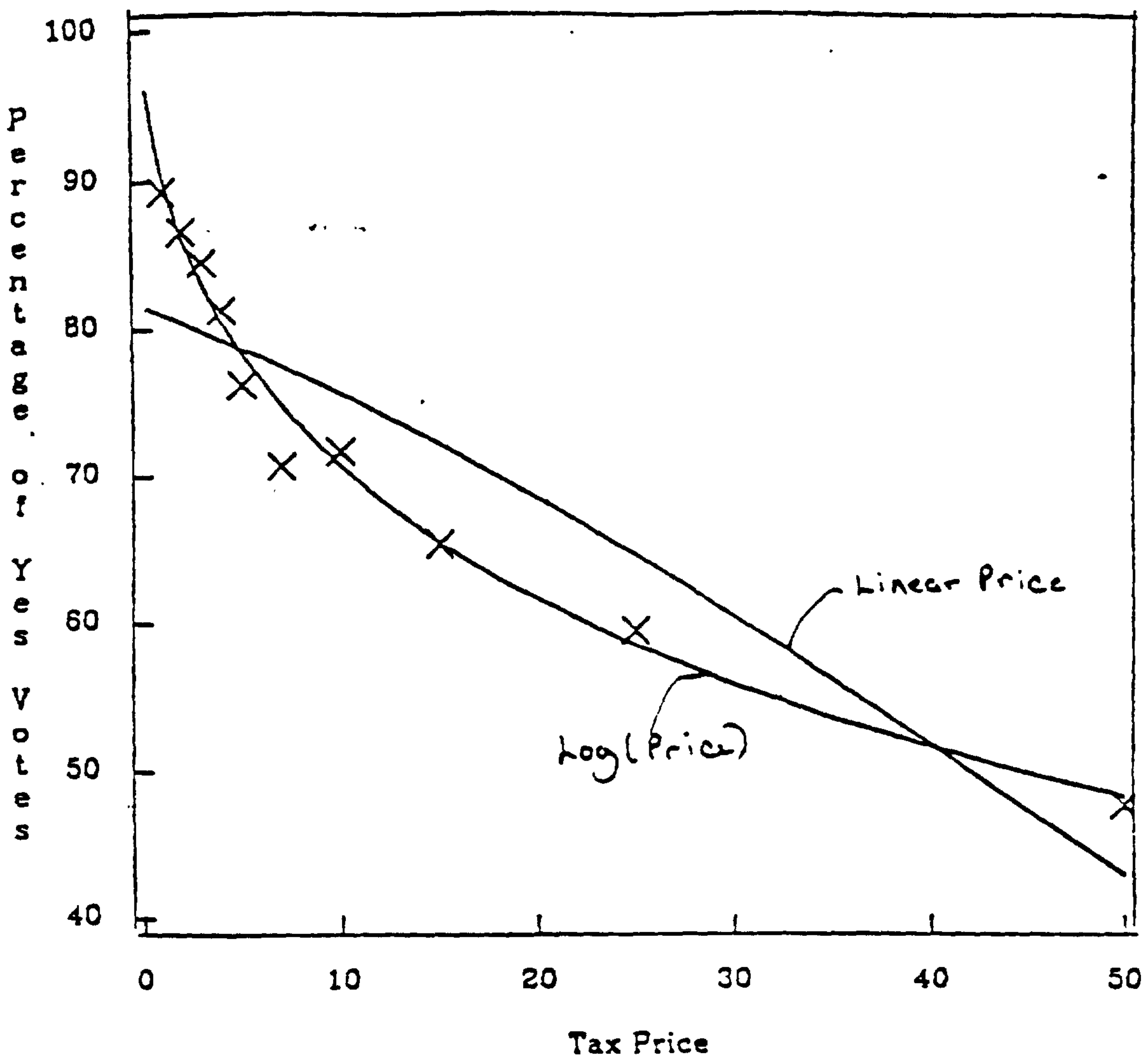
The results presented below are for N*** since this is the correct definition of effective sample size. The results with N* and N** are given in footnotes and are very similar. Ln(price) as the stimulus variable provides a better fit to the data than the model with linear price. For the logit equations, the price model has a X^2 of 69.74 (df=8) while the ln(price) model has a X^2 of 8.38 (df=8).[33] Since the scale of the binomial distribution is one, the expected value of X^2, if the model is appropriate, is 8, indicating that the ln(price) model fits well while the linear price model does not.

This can best be seen graphically. Figure 3-1 is graphed in the manner of a bioassay to emphasize the nature of tax price price as the stimulus variable under the researcher’s control. The ln(price) model fits the data everywhere while the linear price model cannot really be said to fit the model well in anything but a very small region.[34] The actual points are marked with small boxes.

33. The X^2 statistics for the N* are 2.36 and 16.47. and 7.31 and 89.76 for the N** models, where the larger number in both cases is for the linear price model.

Figure 3-1

TAX PRICE BY PERCENTAGE OF YES VOTES



CURVES FIT BY TWO LOGIST REGRESSION MODELS AND ACTUAL OBSERVATIONS

The estimated equation is

$$E[\ln(p_j)/(1 - p_j)] = 2.256 - .6010*\ln(\text{price}) \,, \qquad (11)$$

where the asymptotic standard error of the intercept term is .0758 and the asymptotic standard error of the ln(price) term is .03757.[35] The tax price . elasticity of voter approval ε_v can be defined here as:

$$\partial L(X_k\beta)/\partial X_k = \frac{\exp(x_k\beta)}{(1+\exp(x_k\beta))} * \beta_k, \qquad (12)$$

where L( ) is the logistic function and x_k is the variable of interest [here ln(price)]. Using the coefficients from EQ. (10) and evaluating EQ. (11) at a tax price of $4, gives a point estimate of ε_v as -.1821. The approximate 95% confidence interval for ε_v is -.1714 to -.1902.[36] Figure 3-2 displays the actual demand curve for the public good represented by Proposition 25 as a function of price in the traditional economic manner with price on the vertical axis and quantity on the horizontal axis. This demand curve looks as if it could have been lifted from one of the standard elementary "principles"

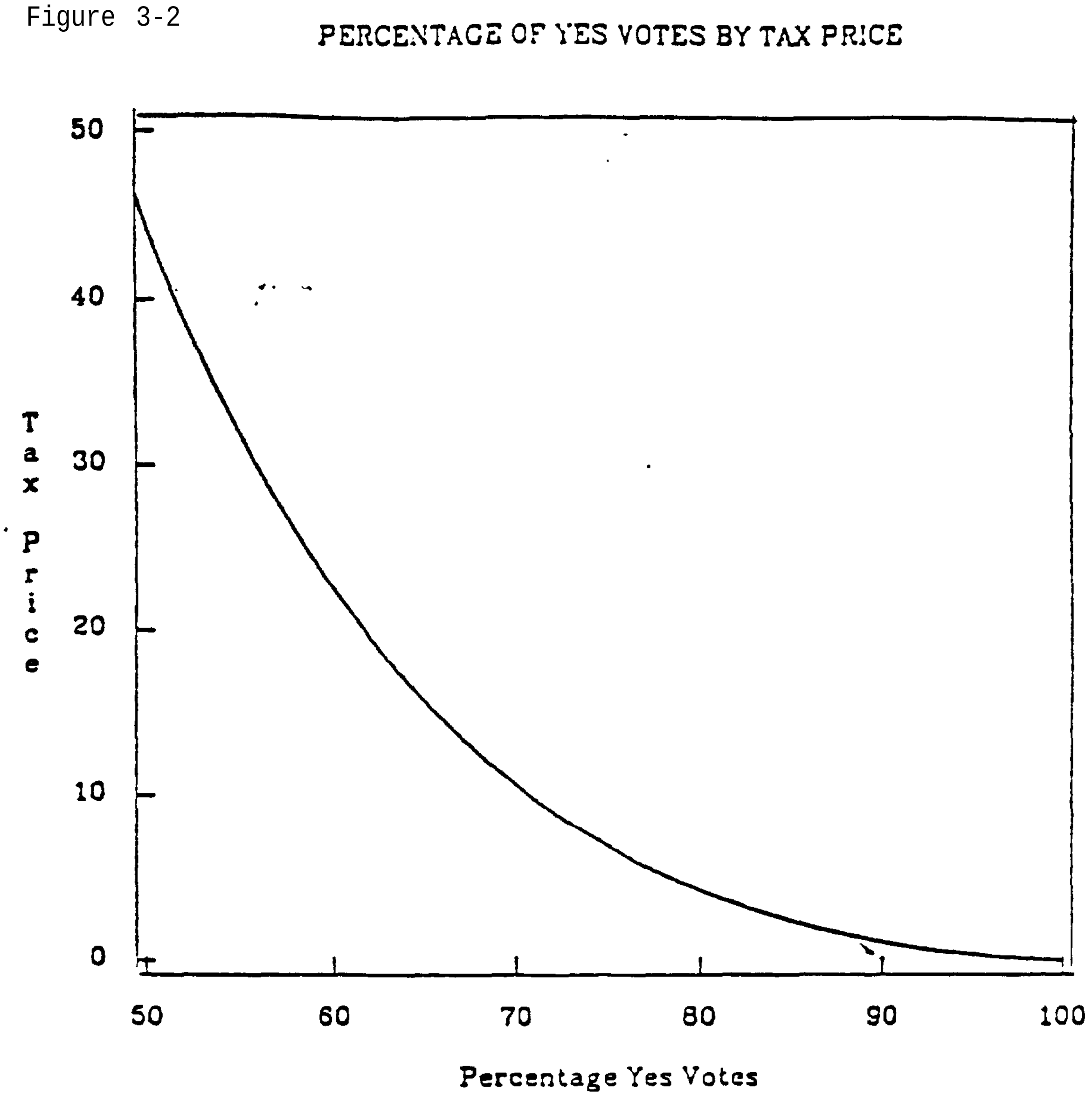
---

34. One of the advantages of this approach with its small number of data points (not observations though) and only one stimulus variable is that bad fits are obvious and desirable transformations easily seen.

35. An estimation of the regression diagnostics suggested by Pregibon (1961) or logistic regression reveal no problems with EQ. (10). For the N* ln(price) model, the coefficients are 2.264 (.1337) and -.5986 (.06058). and for N**, 2.236 (.0625) and -.5888 (.03031) where the numbers in parentheses are the accompanying asymptotic standard errors.

36. For the N* log(price) model, the upper 95% bound, the point estimate, and lower 95% bound for ε_v are [-.1625, - .1818, -.1939] while for N** they are (-.1717, -.1805, -.1876]. It is now clear that the choice of the definition of the n_j is not tremendously important as ε_v is estimated far more precisely than almost any estimate we can find for a public good in the literature. The role of non-sampling survey errors is undoubted more important than any remaining sampling variation at this point.

Figure 3-2


PERCENTAGE OF YES VOTES BY TAX PRICE
Tax Price
50
40
30
20
10
0
50
60
70
80
90
100
Percentage Yes Votes

or public economics texts. It is important to note how we were able to estimate such a curve. Following from those principles, the demand for a good is a function of price, income and taste. At each of the different tax prices, we have a random sample of voters from the same population. Thus, income and taste variables are held constant. The only thing which has changed is the tax price. Because of the controlled application of tax price as a stimulus, there are none of the usual separability problems associated with estimating demand functions. There are also none of the usual simultaneity problems. The work presented here can be taken as a demonstration that the demand curves from micro principles exist and are nicely behaved when the conditions we were first taught can be observed.

Characteristics Associated with Yes on VOTE4

Past work on the determinants of willingness to pay for water quality (Carson and Mitchell, 1986) suggests the following demographic or attitudinal variables (self-identification as an environmentalist, concern over water pollution, participation in water-based recreation, and income) are associated with higher willingness to pay for water quality improvements. Other researchers have occasionally reported age, race or sex correlations. Since there are often partisan differences on referenda, we also selected political ideology and presidential preferences as additional possible correlates.

The hypothesis of independence can be tested using the χ^2 test described above. Before doing this, we dichotomize VOTE4 so that it equals one if the respondent was willing to pay $4 and equals zero if the respondent did not indicate a willingness to pay $4.

For the 2x2 table of preference for Reagan by VOTE4, we have a χ^2 value of .124 (df=1) indicating that the null hypothesis of no association cannot be rejected. The 2x2 table of Republican/Non-Republican [PARTY], shows some support for the hypothesis of that partisan differences influence VOTE4 since the χ^2 value is 4.28 (df=1) which has a p-value of .04.

Income is measured in this survey using the California Poll's standard series of categories. Unfortunately over 30% of the respondents are in the highest income category (total household income greater than or equal to 40 thousand dollars) with most of the detail available for low income households. At a $4 tax price, the income elasticity appears to be close to zero. This is perhaps not too surprising since the cost of the referendum at $4 is not burdensome. As one increases the tax price from $4 to $50 the income elasticity appears to increase but still remains fairly small.[37] We also find no race or sex effects.

For the water based recreation variable [WATUSE], we find that the hypothesis of independence is rejected [χ^2 = 4.91 (df=1; p-value=.03)]. We can also reject the hypothesis of independence with respect to an economic-environmental tradeoff variable, ECEV, [χ^2 = 33.8 (df=4; p-value=.001)], for confidence in the state legislature to provide environmental protection, CONFLEG, [χ_2 = 28.18 (df=3; p-value=.001) and for self-identification as an environmentalist, ENVIST, [χ^2 = 21.38, (df=2; p-value= .001)].[38] WATUSE takes

37. We hesitate to speculate on its exact functional form due to data problems and estimation problems, and the fact that our experimental design was not intended for this purpose. A typical CV survey uses modified "standard" demographic question in order to obtain more detail, particularly with regard to income.

38. WATUSE takes two values, 1 equal to "engage in water-based recreation" and 0 equal to "do not engage in water-based recreation". ECEV is a five point scale with 1 equal to "very pro-environmental with respect to economic trade-offs" and 5 equal to "very pro-economic growth with respect to environmental trade-offs." CONFELG is a four point scale with 1 equal to "great confidence in the state legislature in providing environmental protection" and 4 equal to "no confidence in the state legislature".

two values, 1 equal to “engage in water-based recreation” and 0 equal to “do not engage in water-based recreation.” ECEV is a five point scale with 1 equal to “very pro-environmental with respect to economic trade-offs” and 5 equal to “very pro-economic growth with respect to environmental trade-offs.” CONFLEG is a four point scale with 1 equal to “great confidence in the state legislature in providing environmental protection” and 4 equal to “no confidence in the state legislature.”

CONCLUSION: KNOWLEDGE AND THE ACT OF VOTING

Perhaps the strongest criticism of willingness to pay estimates obtained using the contingent valuation has been put forth by Bishop, Heberlein, and Kealy (1983) and Freeman (1986). In the words of Freeman, “contingent valuation works best where we need it least.” By this, Freeman was referring to the fact that people are most likely to give reasoned and informed willingness-to-pay answers about public goods and levels of public goods with which they have had the most experience. In particular, he refers to recreational demand where methods (e.g., travel cost analysis [Clawson and Knetsch, 1966]) based on complementarity with marketed goods already exist for benefits estimation. Bishop, Heberlein and Kealy (1983) explore this theme further, noting the long process of comparison and consultation a consumer is likely to go through before purchasing a large durable item such as an automobile.

This is undoubtably true of purchases such as automobiles. We put forth another model, though, and that is the actions for a mental process which an individual goes through before casting a vote in an election. It is important to note that Bishop, Heberlein, and Kealy and Freeman have raised their criticisms in response to the hypothetical nature of contingent valuation surveys and the fact that responses given in a contingent valuation survey are not binding.

There is a vast literature in political science which suggests that people are not tremendously well informed when they cast votes.[39] Only a few days before the election, less than 50% of the likely voters were aware that Proposition 25 was on the ballot. This was in spite of media endorsements and a detailed voter's pamphlet from the Secretary of State's office. Magleby (1984) cites research showing that fewer than 30% o California voters read the voters pamphlet in spite of the fact that it is free and is delivered to their homes. The description of the propositions are complicated, full of legal jargon, and way above the reading level of many of the voters. And yet, people vote, making decisions which have a tremendous influence on their lives.

We assert that the half hour to an hour contingent valuation survey interview with its tremendous detail on what the problem is, what the respondent is purchasing, what the alternatives are, and how the solutions can be implemented results in responses that are far more informed that those of rational voters. The method we have proposed here is closer to the actual act of voting but certainly is not the result of informed voters as is a contingent valuation survey. It is much cheaper than a full blown contingent valuation survey. The survey questions on which the estimate of ε_v is based cost less than ten thousand dollars. One thousand 45 minute contingent valuation interviews cost between between seventy-five and one hundred thousand dollars.

---

39. They do, however, vote in a rational fashion and in a manner consistent with their preferences on that issue. Voters do, however, hold contradictory positions which they have not confronted and are often confused about the positions of candidates on minor issues and of the
(Footnote continued)

A contingent valuation survey provides much more information. It provides direct estimates of the Hicksian compensating surplus as well as the actual amounts at which a respondent would change from a yes to a no vote. Typically, one can also explore the provision of different quantities of the public good in a contingent valuation survey. For many purposes, estimates of ε_v prices are sufficient.

The study we have presented in this chapter also provides indirect support for the validity of our national water benefits study as it shows that surveys can predict referenda on water bond issues. Although we did not explicitly invoke the referendum model in the NWBS, our scenario in that study is consistent with the model. Future CV surveys, even of national level programs, should be able to profitably frame the elicitation questions to capture referendum-like choices whether or not they adopt the particular discrete choice framework which we describe here.

39(continued)
exact implications of legislation. See Mueller (1969), Bendict (1980). Crosby, Gill, and Taylor (1981), and Magleby (1984).